

Joint Inference of Misaligned Irregular Time Series with Application to Greenland Ice Core Data

Thinh K. Doan¹, Andrew C. Parnell², and John Haslett¹

¹School of Computer Science & Statistics, Trinity College Dublin, Ireland

²School of Mathematical Sciences, University College Dublin, Ireland

Correspondence to: T. K. Doan (tdoan@tcd.ie)

Abstract. Ice cores provide insight into the past climate over many millennia. Due to ice compaction, the raw data for any single core are irregular in time. Multiple cores have different irregularities; jointly these series are misaligned. After processing, such data are made available to researchers as regular time series: a data product. Typically, these cores are independently processed. In this paper, we consider a fast Bayesian method for the joint processing of multiple irregular series. This is shown to be more efficient. Further, our approach permits a realistic modelling of the impact of the multiple sources of uncertainty. The methodology is illustrated with the analysis of a pair of ice cores (GISP2 and GRIP). Our data products, in the form of marginal posterior distributions on an arbitrary temporal grid, are finite Gaussian mixtures. We can also produce sample paths from the joint posterior distribution to study non-linear functionals of interest. More generally, the concept of joint analysis via hierarchical Gaussian process model can be widely extended as the models used can be viewed within the larger context of continuous space-time processes.

1 Introduction

Ice cores play an important role in revealing Earth's climate history via the analysis of their chemical composition. Data from ice cores are typically available in two forms: "raw" (irregular in time) and "data products" (regularly-spaced in time, i.e. "gridded"). Often the latter are used as input for climate models to analyse past climate change. Alternatively, being gridded, researchers can simply combine one series with other data series, similarly gridded, representing other aspects of climate. Such data products are pre-processed from raw data using a variety of techniques; from simple running averages (Stuiver and Grootes, 2000; Thomas et al., 2007) to complex parametric smoothing (Peavoy and

Franzke, 2010; Nieto-Barajas and Sinha, 2014). These imputation techniques - broadly interpolators or smoothers - are typically applied one core at a time.

This paper proposes a joint statistical model for processing multiple misaligned ice core time series data. Like others in the wider field of palaeoclimate reconstruction (e.g. Haslett et al., 2006; Tingley et al., 2012), our procedure is based on hierarchical Gaussian processes approached through Bayesian inference. Our use of a joint approach is more efficient than its independent alternative as it permits a sharing of strengths across multiple cores, with each core complementing the other in the information needed to learn about the true underlying latent process. A pair of ice cores drilled in Greenland are used to illustrate our proposed framework. Our primary goal is to efficiently take into account multiple sources of information to produce interpretable imputations. A secondary goal is to encourage the use of sample paths of such processes as a richer type of "product".

We introduce and discuss the raw data and their associated uncertainties in Sect. 2. In Sect. 3 we describe our hierarchical Bayesian stochastic process model, and discuss an efficient procedure for inference of the underlying latent process. We report and discuss some results in Sect. 4. One of the advantages of our approach is that researchers can use process sample paths to study non-linear functionals of partially observed processes. As typically published, classical interpolants and smoothers do not facilitate such studies, for they do not typically make available the joint uncertainty associated with the product, a point made pithily by McShane and Wyner (2011) in their critical study of surface temperature reconstructions in the last millennium. Details of the so-called "8.2ka event" (Thomas et al., 2007, e.g.), an abrupt climate change period, provide specific examples. For instance, what precision do such data permit in the timing of this event? In

Sect. 5, we demonstrate that sample paths from our model are useful to answer this question. In Sect. 6, we compare various aspects of the performance of our proposed model against alternative specifications.

Whilst our motivating example stems from climate research, multivariate data with different temporal irregularities are a common feature in many contemporary applications. For example, the ability to combine outputs at different levels of accuracy is crucial to the understanding of processes being studied through potentially expensive computer experimentation. A useful approach in such applications is to combine results from many, cheap but low-resolution experiments with those from a few, expensive but high-resolution experiments, by linking the data via another layer of modelling (Qian and Wu, 2008). In medical applications, one recommendation to overcome issues with misalignment is to align the times before further modelling (Cismondi et al., 2013). For more examples of misaligned time series and associated methods, see Cismondi et al. (2011) and Eckner (2012) and the references therein.

From a theoretical perspective, we view misaligned time series as a special case of spatial misalignment in spatial statistics (Gelfand et al., 2001; Wikle and Berliner, 2005). More generally, our methodological proposal follows the "change of support problem" for data that are indexed in both space and time. This is a generic term that describes methods to combine multiple data sets and make prediction at scales or resolutions which may be different to those of the observations. In this work, we focus on the case where the misaligned observations are at points in time and seek prediction of underlying processes at arbitrary new time points, typically on a regular grid. We offer our perspectives on possible directions of future theoretical and empirical research in Sect. 7.

2 Greenland ice core data

Climate indicators such as tree rings, laminated lake sediments and ice cores are often used as a guide to past climatic conditions (Tingley et al., 2012). This paper focuses on the stable isotope ratios of oxygen in ice cores that are linked to past temperature. The process by which an ice core represents temperature is based on the ease with which two particular types of isotopes in water evaporate from the ocean and condense as snow. We refer interested readers to Jouzel et al. (1997) for a more detailed account of this connection. The measurement is $\delta^{18}O$, recorded as a function of depth. At least for the Holocene period considered in our work, the depth information is transformed to age by counting annual layers in chemicals that show a seasonal cycle (Rasmussen et al., 2006).

We obtain data from the National Snow and Ice Data Center, University of Colorado at Boulder, and the World Data Center for Paleoclimatology, National Geophysical

Data Center, Boulder, Colorado (<http://www.ncdc.noaa.gov/paleo/icecore/greenland/>). In particular, we use the datasets from the United State's Greenland Ice Sheet Project Two (Stuiver, 1999, GISP2) and Europe's Greenland Ice Core Project programs (Johnsen, 1999, GRIP). These are the results from drilling through the Greenland Ice Sheet to recover ice records over 3 km deep. The data sets are presented in several versions; we use a raw dataset in which the cores were cut into sections of equal lengths; 55cm and 200cm for GRIP and GISP2 respectively. A data point is the averaged $\delta^{18}O$ measurement defined at a time point centred at each section of a core. Hence each section of the core can be viewed as the support of a measurement. Even though the sections are equal in length, the ages are irregular. Thus, jointly, multiple cores with different irregularities are temporally misaligned.

We consider the Holocene, a period of relatively stable climate between 0 - 11.5 thousand calendar year Before Present (k cal yr BP), where Present is 1950. $\delta^{18}O$ and date records for all consecutive sections are presented in Fig. 1, along with boxplots of the age increments. Note that we have omitted an age difference value of 80.6 yr between roughly 1320 and 1400 cal yr BP to focus on other significant features of this plot. We see different irregularities in the age increments. Moreover, we note that the ratio of their median values is roughly the same as the ratio in the lengths of the sections, i.e. $\frac{55}{200}$. We discuss aspects of this in App. A.

The empirical semivariograms in Fig. 2 suggest that each series may be adequately modelled via a linear variogram. We defer the discussion of variogram modelling to Sect. 3.1. For the purpose of exploration, we use the R package geoR (Ribeiro Jr and Diggle, 2001) to estimate the parameters of the semivariogram models using the weighted least square method, and compute the total variability of each increment of the $\delta^{18}O$ measurements. We then create QQ plots of the standardised samples as presented in Fig. 3. Apart from two unusual values from each core, these plots indicate the suitability of the Gaussian assumption for the increments. Upon further investigation, we found that the unusual values correspond to neighbouring pairs of measurements with very large negative difference in the $\delta^{18}O$ values.

3 Models and inference

In this section we outline our notation and describe our model for misaligned time series data. Subsequently we show how to perform fast inference on this model without resort to Markov chain Monte Carlo methods, and describe an algorithm for efficient imputation of the latent process onto a grid.

3.1 Hierarchical model

To simplify the narrative in the rest of this article we introduce some notation. Let $y(t_{c,i}); i = 1 \dots n_c$ denote the observed values of $\delta^{18}O$ in core c at time t_i ; as a vector,

we write \mathbf{y}_c . We label the cores as $c = 1, \dots, m$. There are $N = \sum_c n_c$ observations in total, and we use \mathbf{y} to denote these.

We consider a hierarchical model comprised of two main layers. At the observation level $z(t_{c,i})$ is an unobserved value of $\delta^{18}O$. When it is observed, the instrumentation is such that

$$y(t_{c,i}) = z(t_{c,i}) + \nu(t_{c,i}) \quad (1)$$

where the terms $\nu(t_{c,i})$ are iid, zero mean, Gaussian random variables with fixed known variance $\sigma_{\nu_c}^2$ corresponding to the instrumentation. For simplification, hereafter we drop the i subscript when discussing latent processes since they are defined for all times.

At a deeper level, we express $z(t_c)$ as a function of latent value $x(t_c)$ via an additive Gaussian model:

$$z(t_c) = x(t_c) + w(t_c) \quad (2)$$

We propose to model $x(t_c)$ as a continuous-time, independent increments process with increment variance such that for each core $\text{Var}[x(t_c) - x(t_c - h)] = v^2|h|$, i.e. having a linear variogram with no nugget effect. This process may be thought of as driven by climate and thus slowly changing. The process $w(t_c)$ reflects micro-scale, annual-level variations and is modelled by a white noise process, independently across cores, having common $\text{Var}[w(t_c)] = \sigma_{w_c}^2$. Consequently we have

$$y(t_c) = x(t_c) + \nu(t_c) + w(t_c) := x(t_c) + \epsilon(t_c) \quad (3)$$

Here $\epsilon(t_c) \stackrel{\text{ind}}{\sim} \mathcal{N}(0, \sigma_{\epsilon_c}^2)$, with $\sigma_{\epsilon_c}^2 = \sigma_{w_c}^2 + \sigma_{\nu_c}^2$, where the core-specific variance component $\sigma_{\epsilon_c}^2$ is the annual-level nugget effect at core c . The term "nugget" here refers to both sources of variation: the noise at time differences much smaller than the sampling interval, and that which is due to uncertainty from data collection (e.g., Chiles and Delfiner, 2012, Chap. 2.4).

In Sect. 2, we argue that the linear variogram is a suitable model for both ice core data sets. More formally each is consistent with a generating process which is the sum of: (i) an underlying independent increments continuous time process with variogram proportional to lag - the slope; and (ii) a white noise process, manifest in the intercept or nugget effect which is dominated by $\sigma_{w_c}^2$. For our application, the measurement processes differ for each core. In particular both involve averaging a section of core. The chief implication of this is that the nugget effects differ in proportion to the time durations associated with the sections. We can reparameterize so that there is only one nugget term σ_{ϵ}^2 for one of the cores, with the other having a different nugget which is a multiple of this value, i.e., $\sigma_{\epsilon_c}^2 = k_c \sigma_{\epsilon}^2$ for a positive known core-specific value k_c based on the ratio of the measurement periods. For GRIP and GISP2, we set this constant to be the ratio of average supports of 0.275. We outline this choice in more detail in App. A.

The x -increments at different cores will be highly correlated if they reflect climate, and are physically close. In vector forms, \mathbf{x} is the multivariate latent process for all cores at all times, and $\mathbf{x}(t)$ represent all the cores at time t , we write

$$\mathbf{x}(t+h) - \mathbf{x}(t) \sim \mathcal{N}(\mathbf{0}, v^2 |h| \mathbf{\Sigma}) \quad (4)$$

where v^2 is the common marginal variance of an increment per unit time, and $\mathbf{\Sigma}$ is an m -by- m matrix which controls the strength of the relationship of data across the cores. The multivariate process in Eq. (4) forms the basic model underlying our joint approach. To complete the hierarchical specification, prior distributions are assigned to the hyperparameters $\boldsymbol{\theta} = (v^2, \sigma_{\epsilon_c}^2, \mathbf{\Sigma})$. We use reference priors on v^2 , $\sigma_{\epsilon_c}^2$ so that, for example, $\pi(v^2) \propto v^{-2}$. We defer the discussion of model choice for $\mathbf{\Sigma}$ to Sect. 6.

3.2 Imputation

Our objective, given \mathbf{y} , is to provide imputed values of $\delta^{18}O$ for all cores on a regular time grid denoted by $\mathbf{t}_g = \{i\Delta; i = 1, \dots, n_g\}$. These values will be proposed below as expected values of suitable random variables, conditional jointly on \mathbf{y} . Our contribution in this paper is to focus on joint conditioning, in contrast to conditioning on data \mathbf{y}_c in a single core, and in the issues of uncertainty that arise when the various parameters above are themselves only available through statistical inference.

There are two choices of random variable for imputation. We could focus on $E[z(t_c)|\mathbf{y}]$ or $E[x(t_c)|\mathbf{y}]$ for each core c and times $t \in \mathbf{t}_g$. Both are legitimate targets of interest; and their computations are equally straightforward. In this paper we choose the latter; see Eq. (9). Given the parameters appropriate to our data, and in particular that $\sigma_{w_c}^2 \gg \sigma_{\nu_c}^2$ there will be little difference between them, except for cases where a time grid effectively coincides with a time of an observation, and for the corresponding core only. The associated variances, $\text{Var}[z(t_c)|\mathbf{y}]$ and $\text{Var}[x(t_c)|\mathbf{y}]$, will however differ; conditional on the parameters this difference will be approximately $\sigma_{w_c}^2$. We return to this in Sect. 4.

3.3 Inference

Our later mathematical derivations are substantially simplified by defining:

1. \mathbf{t}_o to be the sorted set of all observed times at all cores, i.e. $\mathbf{t}_o = \{t_{c,1}, t_{c,2}, \dots\} = \{t_1, t_2, \dots, t_n\}$ where n is the total number of unique times across all cores.
2. $\mathbf{y}_o(t_{c,i}); c = 1, \dots, m$ whose values coincide with the $y(t_{c,i})$ values when core c has an observation at t_i and whose values are missing otherwise; there is typically one such core for each t_i . Let \mathbf{y}_o denote the set of all such vectors.

This is only a notational trick; the vectors \mathbf{y} (length N) and \mathbf{y}_o (length mn) contain the same information. The objective

is to use this information, together with a hierarchical model of the continuous time multivariate stochastic process \mathbf{x} to impute its values on \mathbf{t}_g . We shall refer to \mathbf{y} and \mathbf{x} defined on \mathbf{t}_g as \mathbf{y}_g and \mathbf{x}_g . Moreover, we let \mathbf{x}_o be the latent multivariate process \mathbf{x} defined on \mathbf{t}_o .

3.3.1 Posterior distributions

In our new notation we can rewrite some of the equations discussed in Sect. 3.1 in their compact matrix form. For instance Eq. (3) can be written as $\mathbf{y}_o|\mathbf{x}_o, \sigma_\epsilon^2 \sim \mathcal{N}(\mathbf{x}_o, \mathbf{Q}_{\epsilon_o}^{-1})$ where \mathbf{Q}_{ϵ_o} is a diagonal precision matrix with entries $(\sigma_\epsilon^2)^{-1}$ and $(k_c \sigma_\epsilon^2)^{-1}$ corresponding to cores at times with data, and zeros where there are no data. This is the notation trick we previously introduced to allow us to write all vectors and matrices in clear order of time and core identity. Note that no computation is ever necessary for components that do not represent data. Similarly, Eq. (4) has the form $\mathbf{x}_o|v^2, \Sigma \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{x_o}^{-1})$ where $\mathbf{Q}_{x_o} = \mathbf{Q}_{t_o} \otimes v^{-2} \Sigma^{-1}$. Here, \mathbf{Q}_{x_o} is a Kronecker product of \mathbf{Q}_{t_o} , the precision matrix of the univariate independent increment process for irregularly-spaced time series data (Rue and Held, 2005, Sect. 3.3), and correlation matrix Σ . We can now write the full posterior distribution as

$$\begin{aligned} \pi(\mathbf{x}_o, \theta|\mathbf{y}_o) &\propto \pi(\mathbf{y}_o|\mathbf{x}_o, \theta) \pi(\mathbf{x}_o|\theta) \pi(\theta) \\ &\propto |\mathbf{Q}_{\epsilon_o}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y}_o - \mathbf{x}_o)^T \mathbf{Q}_{\epsilon_o} (\mathbf{y}_o - \mathbf{x}_o)\right) \\ &\quad \times |\mathbf{Q}_{x_o}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{x}_o^T \mathbf{Q}_{x_o} \mathbf{x}_o\right) \pi(\theta) \\ &\propto \mathcal{N}\left(\mathbf{x}_o; \boldsymbol{\mu}_{x_o|\mathbf{y}_o}, \mathbf{Q}_{x_o|\mathbf{y}_o}^{-1}\right) |\mathbf{Q}_{x_o|\mathbf{y}_o}|^{-\frac{1}{2}} |\mathbf{Q}_{\epsilon_o}|^{\frac{1}{2}} \\ &\quad \times |\mathbf{Q}_{x_o}|^{\frac{1}{2}} \exp\left(\frac{1}{2}\mathbf{y}_o^T \mathbf{Q}_{\epsilon_o} (\boldsymbol{\mu}_{x_o|\mathbf{y}_o} - \mathbf{y}_o)\right) \pi(\theta) \end{aligned} \quad (5)$$

Here $|\mathbf{Q}_{x_o}| = |v^2 \Sigma|^{-(n-1)}$, $\mathbf{Q}_{x_o|\mathbf{y}_o} = \mathbf{Q}_{\epsilon_o} + \mathbf{Q}_{x_o}$, and $\boldsymbol{\mu}_{x_o|\mathbf{y}_o}$ is the solution to $\mathbf{Q}_{x_o|\mathbf{y}_o} \boldsymbol{\mu}_{x_o|\mathbf{y}_o} = \mathbf{Q}_{\epsilon_o} \mathbf{y}_o$. Thus \mathbf{x}_o can be analytically integrated out of $\pi(\mathbf{x}_o, \theta|\mathbf{y}_o)$, yielding $\pi(\theta|\mathbf{y}_o)$, before further computation. Thus, our inference procedure comprises two separate stages as follows.

3.3.2 Inference stage 1: hyperparameters

Initially, we focus on $\pi(\theta|\mathbf{y}_o)$, which the same as $\pi(\theta|\mathbf{y})$. We begin with an optimization routine on the logarithm of $\pi(\theta|\mathbf{y}_o)$ to locate the mode $\hat{\theta}$ and the Hessian matrix evaluated at $\hat{\theta}$; the latter is asymptotically the precision matrix for $\hat{\theta}$. These information will be used as a guide to explore the parameter space of $\pi(\theta|\mathbf{y}_o)$ via a search strategy proposed by Rue et al. (2009, Sect. 3.1). We approximate the continuous distribution for θ , namely $\pi(\theta); \theta \in \Theta$ by the discrete distribution $\pi(\theta_j); \theta_j \in \Theta_J; j = 1, \dots, J$. Θ will be used as a numerical approximation to analytical integrations. For instance, the normalising constant can be dis-

cretely evaluated when there are few hyperparameters, i.e. $\pi(\mathbf{y}_o) = \int_{\theta \in \Theta} \pi(\mathbf{y}_o|\theta) \pi(\theta) d\theta \approx \sum_{\theta_j \in \Theta_J} \pi(\mathbf{y}_o|\theta_j) \pi(\theta_j) \Delta\theta_j$

3.3.3 Inference stage 2: imputation

Our next goal is to derive the marginal distribution of \mathbf{x}_g given all observations. We repeat the aforementioned notational trick by letting the star notation denote the processes defined at both the unique (and sorted) observed times and grid, i.e. $\mathbf{y}_* = (\mathbf{y}_o^T, \mathbf{y}_g^T)^T$ and $\mathbf{x}_* = (\mathbf{x}_o^T, \mathbf{x}_g^T)^T$. The problem becomes that of evaluating

$$\pi(\mathbf{x}_*|\mathbf{y}_*) = \int_{\Theta} \pi(\mathbf{x}_*|\mathbf{y}_*, \theta) \pi(\theta|\mathbf{y}_*) d\theta \quad (6)$$

Derivation of the first quantity in the above integrand is, again, by completing the quadratic form as in Eq. (3). The second quantity is effectively $\pi(\theta|\mathbf{y})$, the joint posterior distribution of the hyperparameters previously derived in Sect. 3.3.2. Importantly, the discrete approximation of the latter renders as summations the integrals that arise in Eq. 6, i.e.

$$\pi(\mathbf{x}_*|\mathbf{y}_*) \approx \sum_{\Theta_J} \mathcal{N}\left(\mathbf{x}_*; \boldsymbol{\mu}_{x_*|\mathbf{y}_*}(\theta_j), \mathbf{Q}_{x_*|\mathbf{y}_*}^{-1}(\theta_j)\right) \pi(\theta_j|\mathbf{y}_*) \Delta\theta_j \quad (7)$$

where $\boldsymbol{\mu}_{x_*|\mathbf{y}_*}$ and $\mathbf{Q}_{x_*|\mathbf{y}_*}$ are the posterior mean and precision of $(\mathbf{x}_*|\mathbf{y}_*, \theta)$. We write $\boldsymbol{\mu}_{x_*|\mathbf{y}_*}(\theta_j)$ and $\mathbf{Q}_{x_*|\mathbf{y}_*}(\theta_j)$ to emphasize that they are the functions of θ_j . Thus, $\pi(\mathbf{x}_*|\mathbf{y}_*)$ is a Gaussian mixture over the posterior samples Θ_J with weights $\alpha_j = \pi(\theta_j|\mathbf{y}) \Delta\theta_j$ already computed in the first inference stage. The joint posterior distribution in Eq. (7) will be useful for the simulation of sample paths. We return to this in Sect. 5.

The marginal posterior distribution of the l^{th} element at a specific core corresponding to a temporal grid of interest may be approximated as finite Gaussian mixture:

$$\pi(x_*^{(l)}|\mathbf{y}_*) \approx \sum_{\Theta_J} \mathcal{N}\left(x_*^{(l)}; \mu_{x_*^{(l)}|\mathbf{y}_*}(\theta_j), \tau_{x_*^{(l)}|\mathbf{y}_*}(\theta_j)\right) \alpha_j \quad (8)$$

where $\mu_{x_*^{(l)}|\mathbf{y}_*}(\theta_j)$, conditioning on the sample value θ_j , is the l^{th} element of the corresponding core identity from $\boldsymbol{\mu}_{x_*|\mathbf{y}_*}$. Similarly, each conditional posterior variance $\tau_{x_*^{(l)}|\mathbf{y}_*}(\theta_j)$ is the l^{th} element of the core identity represented in the diagonal of the covariance matrix. In fact the latter can be computed efficiently from the precision matrix $\mathbf{Q}_{x_*|\mathbf{y}_*}$ without having to perform matrix inversion (Rue and Martino, 2007, Sect. 2). Such predictive distributions, in general non-Gaussian, are the source of any imputations proposed in Sect. 3.2. For example, an imputed value and its variance corresponding to the distribution in Eq. (8) can be computed

as

$$E(x_*^{(l)} | \mathbf{y}_*) \approx \sum_{\Theta_J} \mu_{x_* | y_*}^{(l)}(\theta_j) \alpha_j \quad (9)$$

$$\begin{aligned} Var(x_*^{(l)} | \mathbf{y}_*) \approx & \sum_{\Theta_J} \left(\mu_{x_* | y_*}^{(l)}(\theta_j) \right)^2 \alpha_j - \left(\sum_{\Theta_J} \mu_{x_* | y_*}^{(l)}(\theta_j) \alpha_j \right)^2 \\ & + \sum_{\Theta_J} \tau_{x_* | y_*}^{(l)}(\theta_j) \alpha_j \end{aligned} \quad (10)$$

Posterior means and variances are two summaries of the posterior marginal distributions. Calculations of other summarised statistics such as posterior modes, quantiles, etc. are also straight forward. Hence, for generality we use interquartile ranges (IQR) to quantify the imputations. We emphasize that although there are many computations with large (sparse) matrices, such as solving equations, Cholesky decomposition, etc. we only have to do this J number of times. The fact that all of the matrices are sparse and can be efficiently stored and computed as band matrices presents a further huge reduction in the computation cost (Rue and Held, 2005).

4 Model fitting results

In this section we apply the model framework and inference procedures presented in Sect. 3 to analyse the Greenland ice core datasets described in Sect. 2. When there are two cores, $m = 2$. We model the cross-correlation matrix in Eq. (4) as

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad (11)$$

where ρ is the correlation coefficient. Since temperature processes at nearby locations are always assumed to have (strongly) positively correlated increments, we set 0.5 and 1 as an acceptable range to be used in its prior. The hyperparameters are $\theta = \{v^2, \rho, \sigma_\epsilon^2\}$, which represent the variance per unit increment of the latent process, cross correlation coefficient and nugget effect (for the GRIP core) respectively. Using the inference procedure outlined in Sect. 3.3, we obtain the discrete approximation to the marginal posterior distributions of θ . For brevity of presentation, we apply a smoother to the discrete distributions and present the smoothed version in Fig. 4. It can be seen that the posterior distribution of ρ peaks at a high value close to 1 which indicates a strong spatial relationship between the increments in two cores. This is not surprising since the physical locations of the cores are nearby and both reflect the historical changes in regional temperature of Greenland.

Our main interest lies in the data product derived from both cores jointly. The 95% credible intervals for x_g at bidecadal time intervals are represented in Fig. 5. We compare our results (labelled as DPH) with the data products of GISP2 and GRIP (labelled respectively as SG and R) as reported in Stuiver and Grootes (2000) and Rasmussen et al.

(2006); both of which are based on averaging the measurements of shorter duration. The set of posterior imputed values from DPH can be seen to be much smoother than both SG and R because the nugget is so much larger. For both SG and R the nugget is assumed to be the measurement error which is set very small; see also (Bojarova and Sundberg, 2010, Sect. 3.3). As we previously discussed in Sect. 3.1, our nugget comprises both the measurement error and micro-scale variation. Our choice in Sect. 3.2 removes the variation due to both sources of variation, resulting in a smoother latent process.

To gain a better understanding of the benefit of joint modelling over independent alternatives, we fit an independent increments model with Gaussian noise to each core separately. Note that we suppress the relationship between two nugget effects as discussed in Sect. 3.1; so that each model has two hyperparameters. We defer to Sect. 6 for a more formal discussion of these separate models. In this section, we examine the IQR of posterior marginals of their latent process x_g , in comparison to those of the joint model. It can be seen from Fig. 6 that the IQR in separate models for each core are always higher than those of the joint model. This result indicates that the joint approach utilises information from both cores more effectively when the relationship between the cores (here measured by ρ) is strong. We anticipate that the benefit of joint inference would be more apparent as the number of correlated cores increases.

Several interesting and informative features can also be seen in Fig. 6. There is a consistently bigger difference in the IQR of the joint model and separate model for GISP2 than that for GRIP. This reflects the difference in the temporal resolution, or the number of available data points. The spikes (e.g. around times 0, 1.34, 8.2 k cal BP in GISP2) are also a direct implication of the gaps between times points. An exception is around the 8.2ka event, where the spikes are more influenced by some abrupt changes in the $\delta^{18}O$ measurements. In general, the majority of the spikes in the IQR of the separate models are sharper than those of the joint model. However, the reverse happens around 1.4 k cal BP in GRIP. This is a direct consequence of the (lack of) data points from the other core (GISP2) at that period. Nevertheless, it remains consistent that this spike in the IQR of the joint model is lower than that from the separate model. Finally, we note that the IQR in all cores - in both the separate and joint setting - slightly increases with time. This occurs because temporal resolution decreases when the cores are sampled over sections of identical lengths.

Although this is not a direct comparison with other methods - for neither standard deviations nor IQR are available - it suggests that these ignore valuable information by treating each core separately.

5 Sample paths and the study of the 8.2ka event

The data product discussed in Sect. 4, comprising gridded values, is conventional. It is arguably better than others – in the sense of reduced IQR by being based on all the data jointly. However, unaccompanied by appropriate measures of uncertainty, such products are of limited value as a basis for serious statistical research. As remarked forcefully by McShane and Wyner (2011), even when accompanied by confidence intervals, as in Fig. 5, the value remains limited. For these are “pointwise confidence intervals and are not confidence curves for the entire *sample path*”. That is to say, although the gridded values are jointly based on all the available observations, the joint conditional uncertainty of the unobserved $\delta^{18}O$ histories is not available. One manifestation is that the $\delta^{18}O$ history, as presented in a time series plot of the gridded values with a confidence band as in Fig. 5, is necessarily much smoother than the true (though unknown) history. One simple consequence is that uncertainties for changes in $\delta^{18}O$ are not available from such figures; for these minimally require covariances. The set of posterior means is in fact an average of all the possible histories that are statistically consistent with the model.

A greater challenge again is posed by ‘events’ such as the 8.2ka event, the sudden reduction in North Atlantic temperature during a period around 8.2 k cal yr BP (Thomas et al., 2007). It is believed to be related to a transient change in the North Atlantic overturning circulation. Consequently, the amount of evaporated water in the ocean that became ice in Greenland is amongst its best sources of evidence. The date corresponding to the local minimum of the averages is not a satisfactory estimator of the time of such event. We use sample paths to illustrate a more satisfactory approach. Formally we focus on the random variables $x_{\min} = \min_t x(t)$ and $t_{\min} = \operatorname{argmin}_t x(t)$, being respectively the minimum value for each core, and the date on which this minimum was achieved. Crucially these are non-linear functions of the latent process $x(t)$. Our objective is to make probabilistic statements about these, conditional on all the data. Sample paths provide the methodology.

For brevity, we use Tab. 2 to illustate an example of two sample paths from one core on a grid of 500 yr between time 7.5 and 9 k cal yr BP. These represent independent realisations of the latent process x on the time grid (x_g) . More specifically, they are simulated from $\pi(x_*|y_*, \theta)$, with hyperparameters θ being a sample from $\pi(\theta|y_*)$; both of which are components of the integral (6). The non-linear functionals of interest can then be directly computed for each of very many sample paths and summarised.

A more complete version of this approach uses 1000 sample paths of the gridded process on a bi-decadal time grid over the period of 7.90 to 8.50 k cal yr BP to study the 8.2ka event. Each corresponds to a different fixed value of the sampled parameters θ , and conditional on θ each sample path is of length 30 and is a draw from a Gaussian distribution. This

time range is chosen to bracket the main event, and to distinguish from the possible long term climate trend (Morrill and Jacobsen, 2005). The 1000 minima of the sample paths are summarised in Fig. 7. Additionally, we estimate the interquartile ranges of the timing of the event from GISP2 to be (8.12, 8.16, 8.18) and GRIP to be (8.12, 8.14, 8.18) k cal yr BP. Our findings are consistent with previous studies reported elsewhere; but no quantification of the uncertainty has previously been attempted; see, for instance, Thomas et al. (2007); Kobashi et al. (2007).

An ensemble of sample paths thus provide a flexible data product in its own right. Indeed in climate reconstruction, this is precisely as proposed by Tingley et al. (2012). The use above for minima is illustrative; any function of the process may be studied, conditional on the data. One function, the conditional mean at time t_g , is of course more efficiently computed from its analytical expression as discussed in Eq. (9).

6 Alternative models

In this section, we discuss the model choice for Σ within the context of our application by comparing it with alternative models and choosing the best among them. To formally measure the benefit of joint modelling we compare it with an approach that ignores cross-correlation between cores; this was informally introduced in section 4. Then, we propose an extension to the current model which assumes varying variances for different cores.

Let M_1 denote the structure for Σ defined in Eq. (11). Recall that this assumes equal variance of increments across the cores. We will compare its performance with alternative models, M_2 and M_3 . Model M_2 ignore the cross-correlation relationship between cores; as discussed in Sect. 4. Hence it effectively comprises of two separate models; one for each core. We have demonstrated that the joint approach utilises information from both cores more effectively than the separate approach. We now define M_3 to have an additional parameter in the covariance structure to allow for varying variance between cores, i.e.

$$\Sigma_a = \begin{pmatrix} 1 & \rho\sqrt{a} \\ \rho\sqrt{a} & a \end{pmatrix} \quad (12)$$

The marginal posterior distribution for parameter a of model M_3 is presented Fig. 8. The mode centres around 1 while the marginal posteriors for other parameters (not shown here) are practically the same as those in model M_1 presented in Fig. 4. These suggest that the common variance assumption of model M_1 is most suitable for this data set. Thus, model M_3 is essentially a more conservative version of M_1 , but such an extension may not be necessary when we take into account the additional computational cost associated with an extra hyperparameter. Finally, Table 1 provides further evidence that the extra parameter can be made redun-

dant, as the Bayesian information criterion (BIC) of Schwarz (1978) suggests that M_1 is superior to M_3 .

7 Conclusions

We have presented a hierarchical Bayesian model to jointly analyse multiple misaligned time series. An important component of our model is the Gaussian Markov assumption based on multivariate independent increments that gives us a natural vehicle for joint modelling. We further derived and implemented a fast algorithm for parameter inference based on this model. We applied our method to a pair of GISP2 and GRIP cores, and created data products that are consistent with our empirical knowledge of the physical climate system. More generally, we demonstrated that the joint approach utilises information from multiple cores more efficiently than one-core-at-a-time alternatives.

To the best of our knowledge, our work is a first attempt at directly addressing the joint behaviour of multiple ice cores in their raw and misaligned form. We offer an approach to create data product in the form of non-Gaussian posterior predictive distributions, which is more flexible than what was previously possible. Additionally, our path sampling formulation allows for studies of a variety of non-linear functionals of partially observed processes. Here it is available at relatively little cost.

Some parameters in our model were formulated according to the respective length of the sections of ice cores. Since we consider the fairly stable period of climate conditions, we have not modelled the non-linear age-depth relationship. Furthermore, we did not take into account the error in the dating. Both of these assumptions are likely to be problematic in studying longer sections than the Holocene. A more realistic approach then is to allow for varying change of support. Within the hierarchical Bayesian framework, it is also conceptually straight forward to incorporate uncertainty in the time scale.

In summary, this paper has been tailored to creating data products from Greenland ice cores. In this respect, we feel that it is a proof-of-concept building on the well-established ground of space-time modelling. There are several ways in which our model can be immediately extended to a wide class of spatio-temporal process. For example, we might investigate the possibility of including heterogeneous variability with respect to time by modifying Eq. (4). This corresponds to the spatial extension of the stochastic volatility model for irregular time series recently discussed by Parnell et al. (2014).

Appendix A: Implication of the change of support theory in the measurement procedure

In this appendix we discuss our treatment for the process variance and nugget via the change of support theory. More

specifically, we examine the change in the theoretical semivariogram when there is a change in the underlying support of the data. Without loss of generality, we suppose that the data process \mathbf{y} can be modelled by a noisy univariate independent increments process with a theoretical linear semivariogram of the form $\hat{\sigma}^2 + \frac{1}{2}\hat{v}^2|h|$ where the nugget $\hat{\sigma}^2$ is the intercept, the process variance \hat{v}^2 is twice the slope value, and h is the time lag. We create the new process $\tilde{\mathbf{y}}$ on a new support by averaging \mathbf{y} at every non-overlapping w 'windows' of time. We are interested in the relationship between the semivariogram of \mathbf{y} and that of $\tilde{\mathbf{y}}$.

If \mathbf{y} is a pure nugget process it can be shown that the semivariogram of the averaged process is $\frac{\hat{\sigma}^2}{w}$ when $|h| \geq w$ and $\frac{\hat{\sigma}^2|h|}{w^2}$ when $|h| < w$. For the independent increments case, this is $\frac{1}{2}\hat{v}^2|h| - \frac{w\hat{v}^2}{6}$ when $|h| \geq w$ and $\frac{\hat{v}^2|h|^2}{2w} - \frac{\hat{v}^2|h|^3}{6w^2}$ when $|h| < w$. We refer to Chiles and Delfiner (2012, Chap. 2.4) for the technical details of the aforementioned results. The only lags available for semivariogram calculation in practice are $|h| \geq w$. Thus, for a process that has both the nugget and independent increments, the implication is twofold. First, \mathbf{y} and $\tilde{\mathbf{y}}$ share the same process variance \hat{v}^2 . Second, their nugget parameters are (approximately) directly proportional. This approximation affects the semivariogram near the origin, where we use the linear function to account for the true cubic. From the analytical expression it can be seen that the accuracy of the approximation depends on the value of w , and the relative difference between $\hat{\sigma}^2$ and \hat{v}^2 .

For our application, we assume that the nugget effect is at an annual level (denoted as σ_{annual}^2), with corresponding $\sigma_{GRIP}^2 = \frac{\sigma_{annual}^2}{w_{GRIP}}$ and $\sigma_{GISP2}^2 = \frac{\sigma_{annual}^2}{w_{GISP2}}$, where factors w 's denote the time support for each series. We can reparameterise in terms of $\sigma_\epsilon^2 = \sigma_{GRIP}^2$, such that $\sigma_{GISP}^2 = \sigma_\epsilon^2 \times \frac{w_{GRIP}}{w_{GISP}} = \sigma_\epsilon^2 \times k$. We set $\frac{55}{200}$ or 0.275 as the value for k , corresponding to their respective length of support. This value is also consistent with the descriptive statistics of the age increments, as represented in Fig. 1(c). The sections, being 55cm and 200cm in lengths respectively for GRIP and GISP2, are negligible compare to the total length of the ice core which is about 3km (approximately 1.6 km of which covers the Holocene period). Thus we feel this is a reasonable approximation.

Appendix B: Model validation

As a final model checking step, we determine whether the parameters in our model are identifiable or not. We do this by simulating model parameters ($v^2, \rho, \sigma_\epsilon^2$) based on the results of the data analysis of Greenland ice core, thence the latent bivariate process \mathbf{x} and consequently artificial data \mathbf{y} . We partially average the sequences so as to match the change of support that occurs in our ice core example. We then fit the model (as described in Sect. 3.3) and determine whether the 50 and 90% posterior intervals contain the true values. We

repeat these steps 1000 times, and count up the proportion of occurrences where the intervals contain the true values. A properly calibrated and identifiable 50% interval should contain the true value 50% of the time, and similarly with the 90% interval.

The results of our simulations are shown in Table B1. As can be seen, the intervals contain slightly fewer than the desired proportion of true values, so our posterior intervals are over-precise. However, this effect appears small, and the model seems generally identifiable.

Acknowledgements. The authors thank professor Eric W. Wolff for his helpful comments on the Greenland ice core dataset.

References

- Bojarova, J. and Sundberg, R.: Non-Gaussian state space models in decomposition of ice core time series in long and short time-scales, *Environmetrics*, 21, 562–587, doi:10.1002/env.1012, 2010.
- Chiles, J.-P. and Delfiner, P.: *Geostatistics: modeling spatial uncertainty*, vol. 497, John Wiley & Sons, 2012.
- Cismondi, F., Fialho, A. S., Vieira, S. M., Sousa, J. M., Reti, S. R., Howell, M. D., and Finkelstein, S. N.: Computational intelligence methods for processing misaligned, unevenly sampled time series containing missing data, in: *Computational Intelligence and Data Mining (CIDM)*, 2011 IEEE Symposium on, pp. 224–231, IEEE, 2011.
- Cismondi, F., Fialho, A. S., Vieira, S. M., Reti, S. R., Sousa, J., and Finkelstein, S. N.: Missing data in medical databases: Impute, delete or classify?, *Artificial intelligence in medicine*, 58, 63–72, 2013.
- Eckner, A.: A framework for the analysis of unevenly spaced time series data, Preprint, 2012.
- Gelfand, A. E., Zhu, L., and Carlin, B. P.: On the change of support problem for spatio-temporal data, *Biostatistics*, 2, 31–45, 2001.
- Haslett, J., Whitley, M., Bhattacharya, S., Salter-Townshend, M., Wilson, S. P., Allen, J. R. M., Huntley, B., and Mitchell, F. J. G.: Bayesian palaeoclimate reconstruction, *J. Roy. Statist. Soc. Ser. A*, 169, 395–438, doi:10.1111/j.1467-985X.2006.00429.x, 2006.
- Johnsen, S. J.: GRIP Oxygen Isotopes, <http://doi.pangaea.de/10.1594/PANGAEA.55091>, 1999.
- Jouzel, J., Alley, R. B., Cuffey, K., Dansgaard, W., Grootes, P., Hoffmann, G., Johnsen, S. J., Koster, R., Peel, D., Shuman, C., et al.: Validity of the temperature reconstruction from water isotopes in ice cores, *Journal of Geophysical Research: Oceans* (1978–2012), 102, 26 471–26 487, 1997.
- Kobashi, T., Severinghaus, J. P., Brook, E. J., Barnola, J.-M., and Grachev, A. M.: Precise timing and characterization of abrupt climate change 8200 years ago from air trapped in polar ice, *Quaternary Science Reviews*, 26, 1212–1222, 2007.
- McShane, B. B. and Wyner, A. J.: A statistical analysis of multiple temperature proxies: are reconstructions of surface temperatures over the last 1000 years reliable?, *Ann. Appl. Stat.*, 5, 5–44, doi:10.1214/10-AOAS398, 2011.
- Morrill, C. and Jacobsen, R. M.: How widespread were climate anomalies 8200 years ago?, *Geophysical Research Letters*, 32, 2005.
- Nieto-Barajas, L. E. and Sinha, T.: Bayesian interpolation of unequally spaced time series, *Stochastic Environmental Research and Risk Assessment*, pp. 1–11, 2014.
- Parnell, A. C., Sweeney, J., Doan, T. K., Salter-Townshend, M., Allen, J. R., Huntley, B., and Haslett, J.: Bayesian inference for palaeoclimate with time uncertainty and stochastic volatility, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2014.
- Peavoy, D. and Franzke, C.: Bayesian analysis of rapid climate change during the last glacial using Greenland δ 18 O data, *Climate of the Past*, 6, 787–794, 2010.
- Qian, P. Z. and Wu, C. J.: Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments, *Technometrics*, 50, 192–204, 2008.
- Rasmussen, S. O., Andersen, K. K., Svensson, A., Steffensen, J. P., Vinther, B. M., Clausen, H. B., Siggaard-Andersen, M.-L., Johnsen, S. J., Larsen, L. B., Dahl-Jensen, D., et al.: A new Greenland ice core chronology for the last glacial termination, *Journal of Geophysical Research: Atmospheres* (1984–2012), 111, 2006.
- Ribeiro Jr, P. J. and Diggle, P. J.: geoR: A package for geostatistical analysis, *R news*, 1, 14–18, 2001.
- Rue, H. and Held, L.: *Gaussian Markov random fields*, vol. 104 of *Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC, Boca Raton, FL, doi:10.1201/9780203492024, theory and applications, 2005.
- Rue, H. and Martino, S.: Approximate Bayesian inference for hierarchical Gaussian Markov random field models, *J. Statist. Plann. Inference*, 137, 3177–3192, doi:10.1016/j.jspi.2006.07.016, 2007.
- Rue, H., Martino, S., and Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 71, 319–392, doi:10.1111/j.1467-9868.2008.00700.x, 2009.
- Schwarz, G.: Estimating the dimension of a model, *Ann. Statist.*, 6, 461–464, 1978.
- Stuiver, M.: GISP2 Bidecadal Oxygen Isotope Data, <http://doi.pangaea.de/10.1594/PANGAEA.55531>, 1999.
- Stuiver, M. and Grootes, P. M.: GISP2 oxygen isotope ratios, *Quaternary Research*, 53, 277–284, 2000.
- Thomas, E. R., Wolff, E. W., Mulvaney, R., Steffensen, J. P., Johnsen, S. J., Arrowsmith, C., White, J. W., Vaughn, B., and Popp, T.: The 8.2 ka event from Greenland ice cores, *Quaternary Science Reviews*, 26, 70–81, 2007.
- Tingley, M. P., Craigmille, P. F., Haran, M., Li, B., Mannshardt, E., and Rajaratnam, B.: Piecing together the past: statistical insights into paleoclimatic reconstructions, *Quaternary Science Reviews*, 35, 1–22, 2012.
- Wikle, C. K. and Berliner, L. M.: Combining information across spatial scales, *Technometrics*, 47, 80–91, doi:10.1198/004017004000000572, 2005.

Model	Covariance structure for two cores	$-2\log L$	Penalty	BIC
M_1	$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$	44998	25	<u>45023</u>
M_3	$\begin{pmatrix} 1 & \rho\sqrt{a} \\ \rho\sqrt{a} & a \end{pmatrix}$	44998	33	45031

Table 1. Values of Bayesian Information Criterion (BIC) obtained from different model setting. The underlined value highlights the model with the best fit.

		Sample 1	Sample 2	...
Hyperparameters		(0.2, 0.8, 0.4)	(0.3, 0.9, 0.5)	...
Time	7.5	-34.3	-34.7	...
	8	-35.0	-34.2	...
	8.5	-34.4	-34.4	...
	9	-34.6	-34.2	...
Minimum		-35.0	-34.7	...
Time of minimum		8	7.5	...

Table 2. Illustration of sample paths from one core on a time grid of 500 years intervals. For each sample, we simulate the hyperparameters from their joint posterior distribution, followed by the sample paths of the latent process \mathbf{x}_g . The non-linear functionals of interest are displayed in the last two rows.

Parameter	Proportion inside 50% CI	Proportion inside 90% CI
\mathbf{x}	51%	89%
v^2	48%	88%
ρ	49%	90%
σ_ϵ^2	48%	90%

Table B1. Performance of the model fitting algorithm. All results were based on 1000 simulation runs.

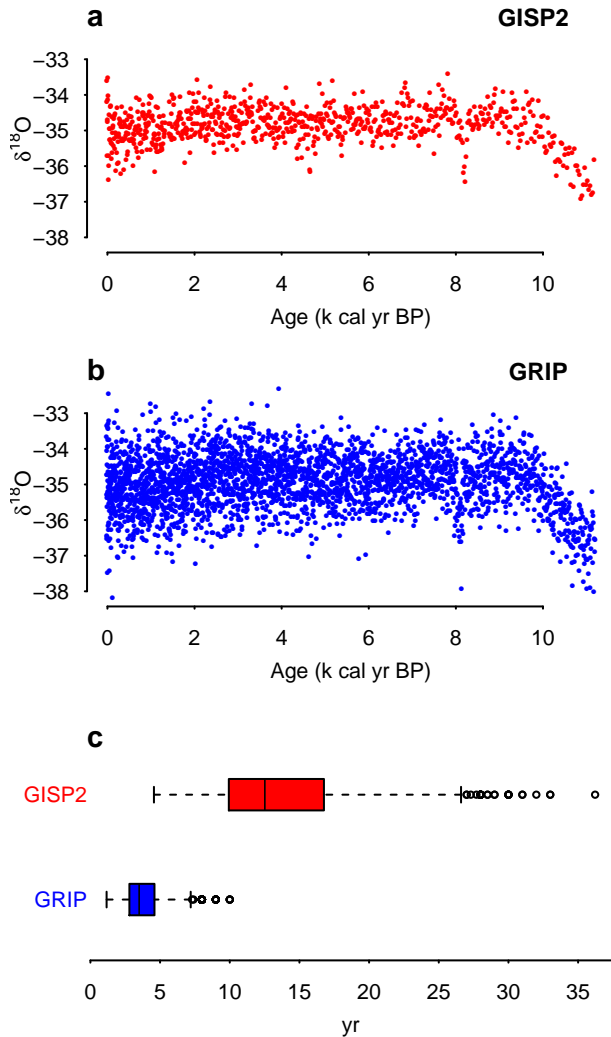


Figure 1. Scatter plots of $\delta^{18}O$ measurements and ages of (a) GISP2 and (b) GRIP. (c) Boxplots of the time increments clearly show different irregularities in the ages. The boxplot for GISP excludes an age difference value of 80.6 yr between roughly 1.32 and 1.4 k cal yr BP. The interquartile range (calculation including the omitted value) are (10.0, 12.5, 16.8) and (2.8, 3.5, 4.6) year for GISP2 and GRIP respectively.

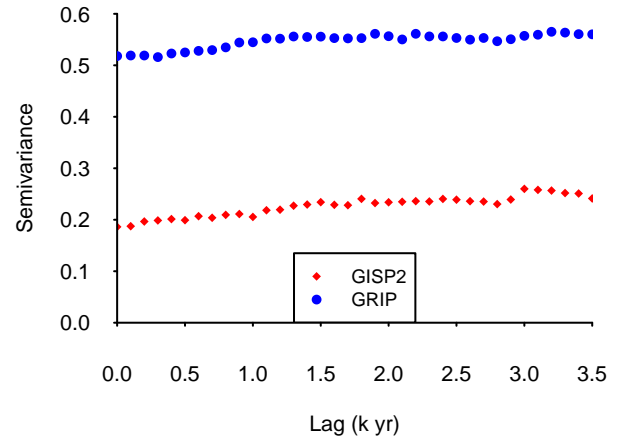


Figure 2. Empirical semivariograms of GISP2 and GRIP. They suggest that the linear variogram is a suitable model for both of the ice core data sets.

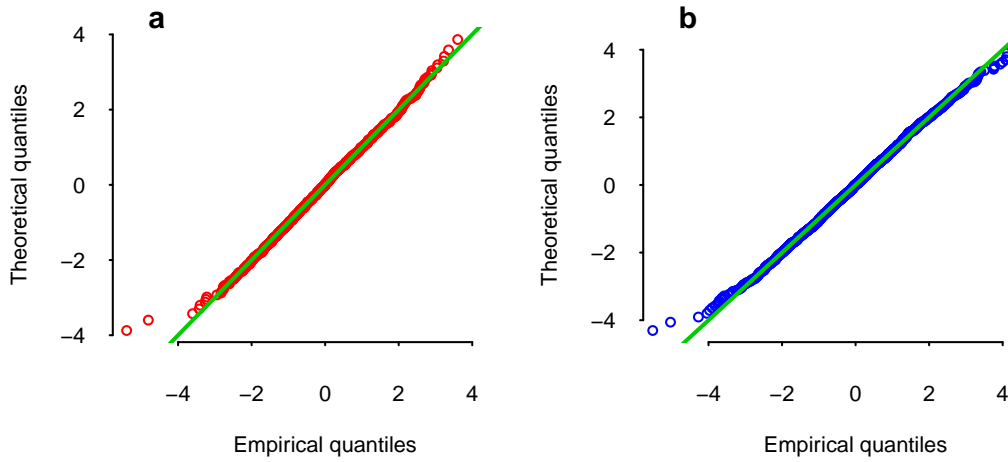


Figure 3. QQ plots of the standardised increments, i.e. the ratio of the first differences in the $\delta^{18}O$ measurements and estimated standard errors of increments for (a) GISP2 and (b) GRIP. The unusual values correspond to consecutive pairs of measurements with very large difference in the $\delta^{18}O$ values.

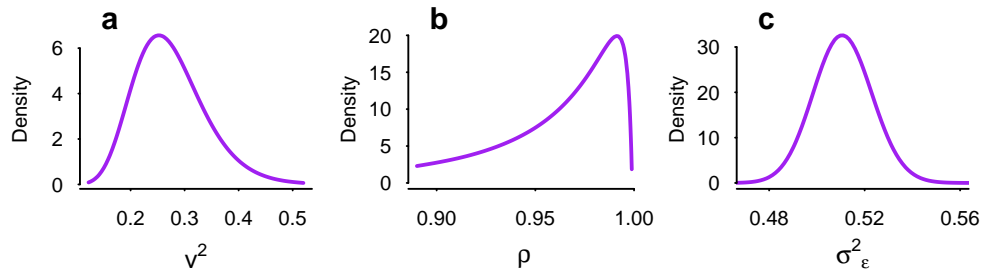


Figure 4. Plots of the smoothed marginal posterior distributions of (a) the variance of the unit increment of x , (b) cross correlation and (c) nugget parameter for GRIP. These are results based on the joint model discussed in Sect. 3.1 applied to both GISP2 and GRIP.

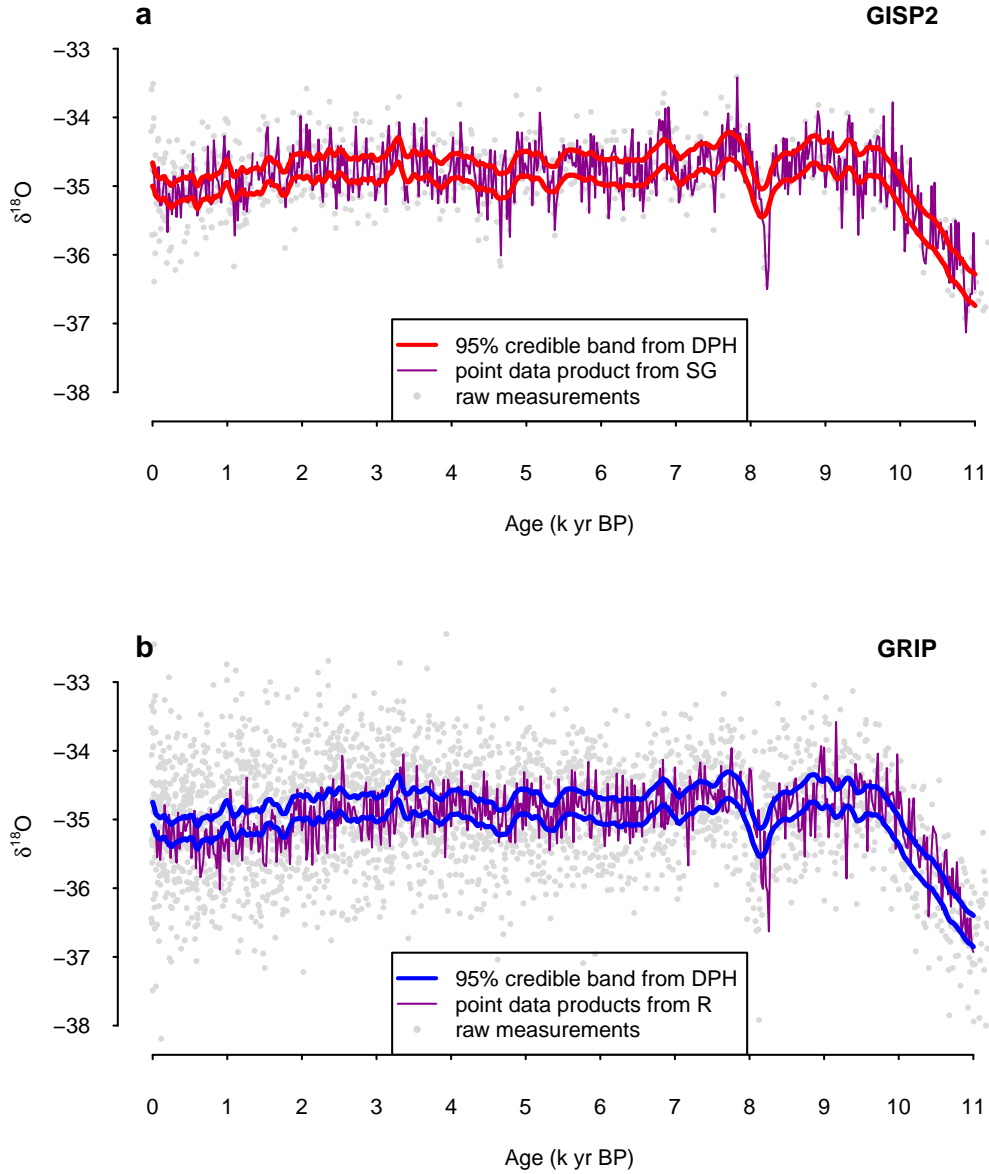


Figure 5. Plots of quantile-based 95% credible intervals of the marginal posterior distributions of process x_g on a bidecadal time grid over the period of 0 to 11 k cal BP conditional on both GISP2 and GRIP. We also show the bidecadal data product from Stuiver and Grootes (2000, SG) and Rasmussen et al. (2006, R).

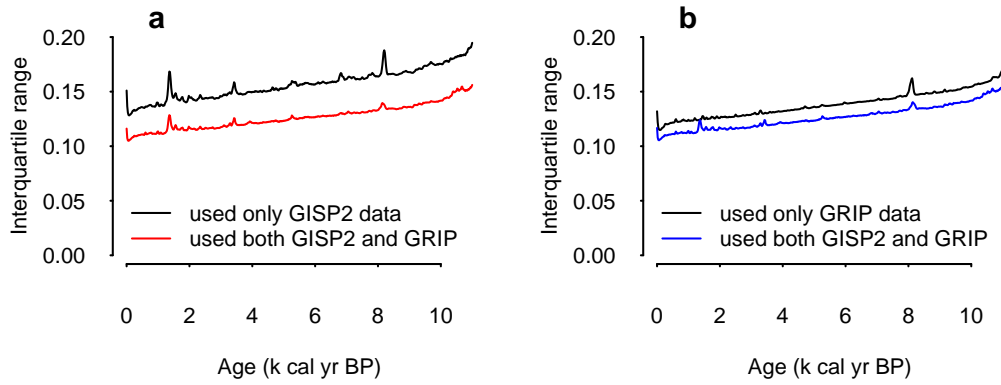


Figure 6. Plots of the interquartile ranges of the marginal posterior distributions of elements of \mathbf{x}_g on a bidecadal time grid over the period of 0 to 11 k cal BP at core (a) GISP2 and (b) GRIP. The main features from these plots are: (i) interquartile ranges of the separate models are always higher than those of the joint model; (ii) the gaps of the differences in (a) are consistently larger than those in (b); (iii) there are several spikes; (iv) a slight tendency for increased IQR further back in time. See the text for a detailed discussion.

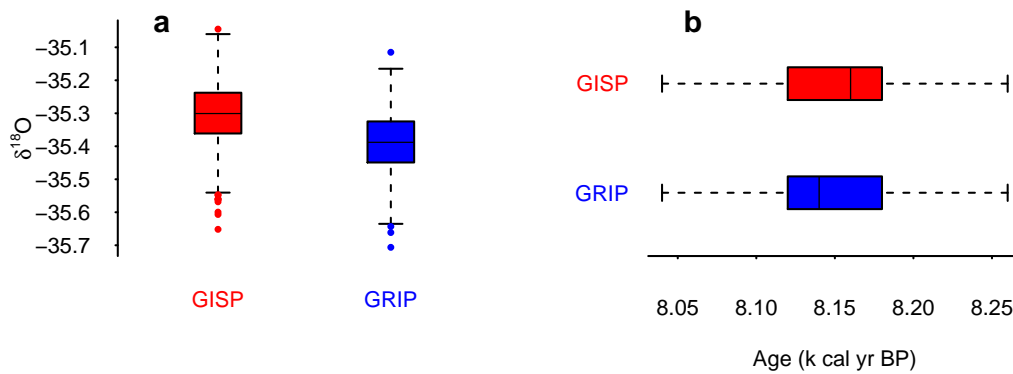


Figure 7. (a) Boxplots of the minima and (b) time of the minima from GISP2 and GRIP over the range of 7.9 to 8.5 k cal yr BP. The interquartile range of the timing of the event from GISP2 are (8.12, 8.16, 8.18) k cal yr BP, and this is (8.12, 8.14, 8.18) for GRIP. All estimates are based on 1000 sample paths.

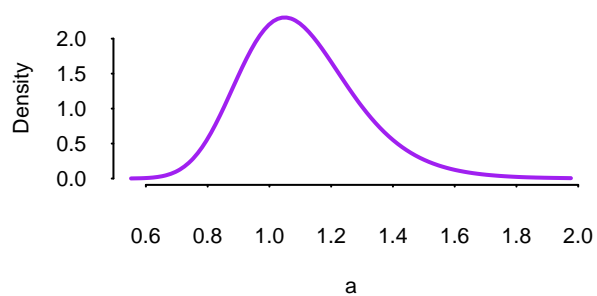


Figure 8. Plot of the smoothed marginal posterior distribution of the parameter a in the covariance matrix of model M_3 described in Sect. 6. The mode is roughly 1 which suggests that a is redundant.